Prognostic value of 18F-FDG PET image-based parameters in oesophageal cancer and impact of tumour delineation methodology.

Mathieu Hatt, Dimitris Visvikis, Nidal M. Albarghach, Florent Tixier, Olivier Pradier, Catherine Cheze-Le Rest

▶ **To cite this version:**

Mathieu Hatt, Dimitris Visvikis, Nidal M. Albarghach, Florent Tixier, Olivier Pradier, et al.. Prognostic value of 18F-FDG PET image-based parameters in oesophageal cancer and impact of tumour delineation methodology.: FDG PET indices for survival prediction. European Journal of Nuclear Medicine and Molecular Imaging, 2011, 38 (7), pp.1191-202. 10.1007/s00259-011-1755-7 . inserm-00574267

HAL Id: inserm-00574267

https://www.hal.inserm.fr/inserm-00574267

Submitted on 22 Jul 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Prognostic value of ^{18}F-FDG PET image-based parameters in esophageal cancer and impact of tumor delineation methodology

Mathieu Hatt[1], Dimitris Visvikis[1], Nidal M. Albarghach[1,3], Florent Tixier[1], Olivier Pradier[1,3], Catherine Cheze-le Rest[1, 2]

[1]INSERM, U650 LaTIM

[2]Academic Department of Nuclear Medicine

[3]Department of Radiotherapy

CHU Morvan, Brest, France

Running title: FDG PET indices for survival prediction

Corresponding author:

Mathieu HATT

LaTIM, INSERM U650

CHU MORVAN

5 avenue Foch

29609 Brest France

Phone:+33298018111

Word count: 5976

Keywords: PET, tumor volume, tumor segmentation, esophageal cancer, survival

ABSTRACT

Background: ^{18}F-FDG PET image-derived parameters, such as standardized uptake value (SUV), functional tumor length (TL) and volume (TV) or total lesion glycolysis (TLG) may be useful for determining prognosis in patients with esophageal carcinoma. The objectives of this work were to investigate the prognostic value of these indices in esophageal cancer patients undergoing combined chemo-radiotherapy treatment and the impact of TV delineation strategies. **Methods**: 45 patients were retrospectively analysed. Tumors were delineated on pretreatment ^{18}F-FDG scans using adaptive threshold and automatic (FLAB) methodologies. $SUV_{max,}$ SUV_{peak}, SUV_{mean}, TL, TV, and TLG were computed. The prognostic value of each parameter for overall survival was investigated using Kaplan-Meier and Cox regression models for univariate and multivariate analyses respectively. **Results**: Large differences were observed between methodologies (from -140% to +50% for TV). SUV measurements were not significant prognostic factors of overall survival, whereas TV, TL and TLG were, irrespective of the segmentation strategy. After multivariate analysis including standard tumor staging, only TV ($p<0.002$) and TL ($p=0.042$) determined using FLAB were independent prognostic factors. **Conclusions**: Whereas no SUV measurement was a significant prognostic factor, TV, TL and TLG were significant prognostic factors of overall survival, irrespective of the delineation methodology. Only functional tumor volume and length derived using FLAB were independent prognostic factors, highlighting the need for accurate and robust PET tumor delineation tools for oncology applications.

1. Introduction

The incidence of esophageal cancer is increasing and despite advances in therapy, the diagnosis still carries a poor prognosis *(1)*. As with all tumors, the outcome for patients is strongly associated with the stage at initial diagnosis. The TNM (Tumor, Node, Metastasis) system currently in use for the staging of esophageal cancer does not take into account non-anatomical factors such as histopathologic type, grade or various biomarkers that may be important determinants of prognosis. The pathologic stage is given by surgery but this procedure is not possible for all patients because it is associated with a high risk of mortality and morbidity. Therefore a non-invasive staging method would be of great interest, and within this context the primary role of ^{18}F-FDG positron emission tomography (PET) in esophageal cancer is the detection of distant metastases *(2-4)*. This modality is also gaining acceptance in esophageal cancer for the assessment of therapy response *(5-6)* or for radiotherapy treatment planning *(7-9)*. Lately, some authors also suggested that different parameters derived from initial ^{18}F-FDG PET images could have a role as independent prognostic factors *(10-14)*. Studied parameters include standardized uptake value (SUV_{max} as the maximum uptake in the primary tumor or in the combined primary and regional area), tumor functional longitudinal length (TL), tumor functional volume (TV), nodal uptake or FDG avid metastases *(10-14)*. Although a few studies have demonstrated the interest of these indices for determining prognosis, there are conflicting results concerning the independent prognostic value of each of these indices. For example, Hyun et al *(12)*, analyzing results from 151 patients with principally squamous cell carcinoma (SCC), have recently suggested that primary tumor SUV_{max} is not an independent prognostic factor, in agreement with other studies *(10,15,16)*. On the other hand, Kato et al *(17)*

based on the analysis of 184 patients with esophageal SCC have shown that SUV_{max} of the primary tumor is an independent prognostic factor of overall survival, in agreement with other studies *(18-20)*. These conflicting results can be potentially caused by differences in the methodology used for the analysis of the PET images. Although SUV_{max} is less sensitive to tumor size, the conflicting results considering its value as an independent prognostic factor may also be due to variability in the tumor sizes considered in the different studies.

Pathological TL has been shown to be an independent prognostic factor in esophageal carcinoma *(21)*. Therefore determining the functional TL in ^{18}F-FDG PET images may be a good surrogate. The use of different thresholds for the determination of the functional TL in the existing studies may be responsible for the conflicting results regarding its value as a predictor of response to chemo-radiotherapy *(11,22)*, while it has been shown as an independent predictor in patients undergoing surgery *(10)*. On the other hand, functional TV may be more representative of overall tumor burden. The value of the functional TV has been only recently explored in a single study of patients with esophageal carcinoma considering a heterogeneous treatment regime (76% and 24% treated by surgery and combined radio-chemotherapy respectively) *(12)*. In this study both the presence of metastases and the TV were found to be independent prognostic factor of patient overall survival. Tumors were delineated based on a three fixed threshold scale depending on the tumor SUV_{max}. Although such an approach may be simple to implement in clinical practice the use of fixed threshold for functional TV determination suffers from multiple shortcomings which have been highlighted in different studies *(23,24)*. In addition, the proposed scale is not universally applicable to the different clinical

settings spanning from the acquisition protocols to the scanning systems used and variable associated image qualities.

Therefore despite early evidence that functional TL and TV may be useful parameters in predicting survival and response to therapy there is a clear need in assessing the influence of the methodology used in obtaining these indices. Finally, the determination of functional TV may allow the calculation of alternative image derived indices such as the total glycolytic lesion index (TLG) (g), defined as the product of the TV (ml) and its associated mean activity (SUV_{mean}) (g/ml) *(25)*, whose value has not as yet been explored in predicting response to therapy or as prognostic factor of survival using ^{18}F-FDG in esophageal carcinoma.

The objective of this study was therefore to retrospectively investigate the prognostic value of ^{18}F-FDG PET in 45 esophageal cancer patients treated with concomitant radio-chemotherapy, considering for the first time in a single study all of the commonly-used PET-derived parameters such as functional TL, TV, SUV measurements (max, peak, mean) and TLG. In addition the impact of different tumor delineation strategies was assessed.

2. Materials and Methods

2.1 Patients

45 patients with a newly diagnosed esophageal cancer treated between 2004 and 2008 with concomitant radio-chemotherapy and without surgery were included in this study. The characteristics of the patients are given in table 1. 41 patients were male (91%), and the mean age at the time of diagnosis was 66±10 years. Most of the tumors were squamous cell carcinoma (73%) and originated from the middle and

lower esophagus (76 %). All patients were referred before treatment for an ^{18}F-FDG PET study as part of a routine procedure for the initial staging in esophageal cancer. The treatment included three courses of 5-fluorouracil/cisplatin and a median radiation dose of 60Gy given in 180cGy daily fractions delivered once daily, 5 days a week for 6-7 weeks. Follow-up data were prospectively collected in a database for further analysis and overall survival was calculated. The current analysis was carried out after an approval by the institutional ethics review board.

2.2 ^{18}F-FDG PET acquisitions

^{18}F-FDG PET studies were carried out prior to the treatment. Patients were instructed to fast for a minimum of 6h before the injection of ^{18}F-FDG. The administered dose was 5 MBq/kg, and static emission images were acquired (2min per bed position) from head to thigh beginning 60 minutes after injection on a Philips GEMINI PET/CT system (Philips Medical Systems, Cleveland, OH USA). In addition to the emission PET scan, a low dose CT scan without IV or oral contrast was acquired for attenuation-correction. Images were reconstructed with the RAMLA 3D algorithm using standard clinical protocol parameters: 2 iterations, relaxation parameter of 0.05, a 5mm 3D Gaussian post-filtering, and a 4x4x4mm^3 voxels grid sampling.

2.3 PET image analysis

The following parameters were extracted from each PET image: SUV_{max}, SUV_{peak} defined as the mean of SUV_{max} and its 26 neighbors, mean SUV within the delineated tumor (SUV_{mean}), functional TL in longitudinal direction, functional TV, and

TLG. SUV_{peak}, usually defined as a 1cm circle or sphere *(26) (w*e used a 3x3x3 cube of 4x4x4 mm^3 voxels which roughly corresponds to the same ROI), was considered in order to investigate the impact of reducing the potential bias in the SUV_{max} measurements as a result of its sensitivity to noise.

Whereas SUV_{max} and SUV_{peak} are independent on the tumor delineation strategy used, TL, TV, SUV_{mean} and the derived TLG were determined on delineations performed using two strategies. First, an adaptive threshold *(23)* using a background region of interest (ROI) manually chosen by two experienced nuclear medicine physicians led to two different results T_{bckgrd}^1 and T_{bckgrd}^2. Observers were instructed to choose the ROI in the mediastinum at a sufficient distance from the lesion to avoid any overlapping. However, they were allowed to choose the size, shape and exact placement of the ROI. Finally the automatic Fuzzy Locally Adaptive Bayesian (FLAB) algorithm *(24,27)* was considered.

2.4 Statistical analysis

All quantitative data were expressed as mean±1 standard deviation (SD) and summary statistics are given in table II.

The correlation between all parameters extracted using the different methodologies was computed with Pearson coefficients. The differences between methodologies were assessed using Bland-Altman analysis *(28)* to define bias as the mean error, and agreements intervals (upper and lower limits) as 1.96 times the SD. Kaplan-Meier methods were used to estimate the survival distributions *(29)*. Overall survival was calculated from the date of initial diagnosis to the date of death or most recent follow-up in case of patients still alive. For each considered parameter, survival curves were generated. The most discriminating threshold value allowing differentiation of the two groups of patients was selected using receiver operating

characteristic (ROC) methodology (30). Prognostic value of each parameter in terms of overall survival was assessed by the log-rank test. The significance of the following factors were tested: age, gender, histology type, T, N, M classifications, AJCC (American Joint Committee on Cancer) stage *(31)*, TL, TV, SUV_{max}, SUV_{peak}, , SUV_{mean}, and TLG. Independent prognostic factors of overall survival were determined using multivariate Cox regression analysis *(32)* by incorporating in the model all parameters that were deemed significant in the univariate analysis. However, the indices obtained by each delineation (T_{bckgrd}^1, T_{bckgrd}^2 and FLAB) were incorporated in the multivariate analysis separately since they were found to be highly correlated (Pearson $r>0.8$, $r^2>0.66$; see results section 3.1). All tests were carried out using MedCalc^{TM} (MedCalc Software, Belgium). *P* values <0.05 were considered statistically significant.

3. Results

All primary lesions were detected by ^{18}F-FDG PET. The intensity of maximum ^{18}F-FDG uptake in the primary lesion was quite high with a normally distributed SUV_{max} of 10±3.8. As expected, SUV_{peak} measurements were comparatively lower (8±3). Measurements related to the dimensions of the tumors were less uniformly distributed than SUV measurements, with a larger SD with respect to the mean. For example the TV(FLAB) was 35±33cm^3.

3.1 Correlation between image derived indices and between methodologies

TL measurements were correlated with TV ($p<0.0001$) although with moderate coefficients ($r=0.69$, 0.58 and 0.6 for FLAB, T_{bckgrd}^1 and T_{bckgrd}^2 respectively). No significant correlation was found between any SUV measurement (SUV_{max}, SUV_{peak},

SUV_{mean}) and TV ($p>0.2$, $r<0.3$), irrespective of the delineation strategy, in line with results of other studies such as Van Heijl et al *(33)*.

All SUV_{mean} measurements derived from TV delineation performed using the two different methodologies considered were highly correlated ($p<0.0001$) with coefficients >0.97. TV ($r>0.82$), TL ($r>0.91$) and TLG ($r>0.95$) results were also highly correlated ($p<0.0001$) for both methodologies.

Despite high correlation coefficients, large differences were observed for several patients between measurements using the two delineation methodologies considered, and between the two users of the same adaptive thresholding. Figure 1(A-B) illustrates such differences. In the case of adaptive thresholding these differences were the result of the two users placing differently the background ROI.

A summary of the Bland-Altman analysis carried out to compare the delineation methods and highlight potential differences is presented in figure 2(C-D) and complete results are given in table III. The largest differences between methodologies were observed for TV with a bias of up to 50% between the adaptative thresholding and FLAB: both users resulted in globally smaller volumes (bias of -50%±50% and -21%±54% for T_{bckd}^{1} and T_{bckd}^{2} respectively). Agreement intervals (upper and lower limits) were large for all parameters and for all comparisons, up to +80% and -140% (see fig.2B). Even between the two users of the same adaptive thresholding method (see fig.2A), mean differences of -30±35% were seen and limits of agreement were large, from -100 to +45%. No significant trend was found regarding the correlation between TV and differences between methodologies ($r<0.2$, $p>0.1$).

Better agreement was observed for TL and SUV_{mean}, however intervals of agreement were large (-50% to -25% lower limit and +20% to +40% upper limit for TL; -80% to -10% lower limit and +10 to +80% upper limit for SUV_{mean}).

Due to the combined effect of TV and SUV_{mean}, TLG differences were in between, with moderate bias but still large agreement intervals (upper and lower limits of -120% to -75% and +40% to +90% respectively).

3.2 Survival analysis

At the time of last follow-up, 10 patients were alive with no evidence of disease, 9 were alive with recurrent esophageal cancer and 26 had died from the disease. With a median follow-up of 60 months (range 9-82), the overall median survival was 15 months. The 1-year and 2-year survival rates were 63% and 34% respectively.

The results of the log-rank analysis of significant parameters for overall survival in univariate analysis are given in table IV. Table V summarizes the prognostic value of all the parameters under investigation in this study.

Age, gender, and histology types were not significant prognostic factors in the univariate analysis. Neither were T and N classifications. In the univariate analysis, the presence of metastases (median survival of 26 months (M0) versus 12 months (M1), p=0.01)) and the clinical AJCC stage ($p<0.001$) were significant prognostic factors.

Although there was a trend observed, neither SUV_{max} nor SUV_{peak} were significant prognostic factors. A SUV_{max} <5 or <8 tend to be a factor of better outcome with a median survival of 14 vs. 7 months (p=0.08) or 21 vs. 13 months (p=0.1) respectively (see fig.3A).

Mean SUVs in the tumor were not significant prognostic factors in univariate analysis. There was however a trend for shorter survival associated with higher SUV_{mean}. For example the median survival reduced by a factor of 2 for patients with a SUV_{mean} higher than 5 (13 months vs. 21 months, p=0.06). This was however observed only when the FLAB methodology was used to define TV, while no similar trend was observed with SUV_{mean} parameters obtained with adaptive thresholding.

Functional TV was a significant prognostic factor of overall survival, whatever methodology was used (p<0.001 using FLAB, and p=0.004 for both T_{bckgrd}^1 and T_{bckgrd}^2, see figure 3(B-C)). In addition, using the TV, and independently of the delineation approach used, allowed to separate our population into 3 groups with significantly different outcome (p=0.002, p=0.02 and p=0.004 for FLAB, T_{bckgrd}^1 and T_{bckgrd}^2 respectively). For instance, volumes defined by FLAB less than $14cm^3$, between 14 and $85cm^3$ or superior to $85cm^3$ were respectively associated with a median survival of 49 (19 patients), 15 (21 patients) and 5.5 (6 patients) months as illustrated in figure 3(D). In figure 4(A-C) three examples of ^{18}F-FDG PET baseline images of patients belonging to each of these three groups are presented.

Functional TL was also a significant prognostic factor with results similar to TV (p=0.01, p=0.02 and p=0.04 for FLAB, T_{bckgrd}^1 and T_{bckgrd}^2 respectively), apart from not being possible to significantly differentiate 3 groups of patients with different outcome, independently of the strategy.

Similarly, TLG was also a significant prognostic factor whatever methodology was used, while it was equally not possible to significantly differentiate three groups. Median overall survival was 10 months for patients with a TLG (FLAB) >180g, and increased to 21 months for patients with a TLG<180g (p=0.01). Similar results were

obtained with adaptive thresholding (20 versus 8 and 20 versus 10 months for T_{bckgrd}^{1} and T_{bckgrd}^{2} respectively).

After multivariate analysis, considering each delineation methodology separately only TV obtained using FLAB and AJCC stage were found to be independent significant prognostic factors (p=0.0017 and 0.0021 for TV and AJCC respectively), whereas only AJCC stage was an independent significant prognostic factor (p<0.002) when considering TV obtained by adaptive thresholding.

Similar results were obtained when replacing TV by TL, with both TL and clinical AJCC staging found to be independent significant prognostic factors in the case of FLAB (p=0.017 and p=0.042 for AJCC stage and TL respectively), whereas in the case of adaptive thresholding only AJCC staging was an independent significant prognostic factor (p=0.0021).

On the other hand, in the case of TLG only the AJCC staging was an independent significant prognostic factor (p<0.002), whatever delineation strategy was considered.

4. Discussion

An accurate staging in esophageal cancer is essential for guiding therapy. The standard conventional modalities are endoscopic ultrasonography and computed tomography even if this combined approach suffers from several shortcomings. ^{18}F-FDG PET is more and more often included in the initial staging because it allows a more accurate disease assessment, especially regarding the detection of distant metastases *(2-4)*. Since no patient underwent surgery in our study, anatomopathology data were not available. Therefore T,N,M classifications and AJCC stages were determined using suboptimal conventional staging and this could explain the poor prognostic value of T or N classification in our population.

As found in our study, ^{18}F-FDG uptake is always present in esophageal cancer if extended at least to submucosa *(34)*. Some authors suggested that the intensity of ^{18}F-FDG uptake could be related to prognosis in esophageal cancer, based on the good correlation existing between hexokinase activity or poor differentiation and tumor uptake *(35)* and also because increasing SUV_{max} values seem to correlate with T classification, which is part of the TNM staging *(36)*.

In our study, SUV measurements were not significant prognostic factors of overall survival. While various cut-off values of SUV_{max} tend to be associated with a poor prognosis, none led to statistically significant differentiation. Swisher et al. reported similar results in a uniform group of highly selected patients with locally advanced esophageal cancer treated by neoadjuvant radiochemotherapy *(37)*. On the other hand, these results could appear in contrast with our previous report *(18),* where we initially reported that a SUV_{max} cut-off value of 9 had an independent prognostic value of overall survival, but this difference may be explained by the different patient characteristics considered in the two studies. We previously considered *(18)* a daily practice population, half of which underwent curative surgery, while we included here only patients with advanced disease exclusively treated by combined radio-chemotherapy.

TL established by pathological examination has been demonstrated to be an independent prognostic factor of long term survival *(21)*. Some authors proposed estimating TL based on ^{18}F-FDG PET images using different thresholds *(38)*. Functional TL has been studied as a predictor of response to neoadjuvant chemoradiotherapy with conflicting results *(11,22)*. In a group of 69 patients with esophageal squamous cell carcinoma undergoing curative surgery, Choi et al. demonstrated that functional TL was an independent prognostic factor *(10)*. However,

one may argue that functional TL is a parameter that does not reflect the real volume of the tumor but only its longitudinal extension and could be therefore considered as only a surrogate of tumor spatial extent. This argument can be supported by the data shown in this work, where only a moderate correlation ($r<0.7$) was found between TV and TL, suggesting that functional TV may be more accurate in assessing actual tumor burden. In our study we also compared the prognostic value of TL with that of TV. Both parameters were found to be significant prognostic factors irrespective of the functional volume delineation strategy. In addition, both TV and TL were independent prognostic factors of survival in the multivariate analysis. However, this result was found to be dependent on the segmentation algorithm, with both parameters being independent survival prognostic factors only when determined using the automatic FLAB segmentation. This may be related to the higher overall accuracy of FLAB with respect to adaptive thresholding for tumor delineation as previously reported *(24,27,39)*. Despite the similar prognostic values of TL and TV, only TV allowed a statistically significant stratification of patients into three groups, irrespective of the segmentation methodology. More specifically, two different cut-off values (85 and 14cm^3) resulted in significant differentiation of two groups among the patients with median overall survival of 5 to 6 vs 20 months ($p=0.0005$) and 49 vs 13 months ($p=0.036$) for 85 and 14cm^3 respectively. Being able to provide such a finer stratification of patient groups could be of value in clinical trials assessing new therapeutic regimes.

SUV_{mean} measured in a volume determined using the different tumor delineation approaches considered was not found to be a prognostic factor of overall survival, although a trend was seen for SUV_{mean} associated with TV defined with FLAB, which

tended to differentiate patients with poor and better prognosis (13 vs 21 months, p=0.06).

A fundamental biological question underlying ^{18}F-FDG PET prognostic value is whether the total volume or the metabolic active portion of the tumor is most important. Intuitively both would seem important and desirable to determine. In our study, both functional TL and TV (representative of the tumor functional spatial extent) were significant prognostic factors compared to SUV_{mean} (representative of the tumor glycolytic metabolism) which was not. Combining both parameters into total lesion glycolysis index (TLG) was a prognostic factor of overall survival whatever methodology was used for tumor delineation. However it was not an independent significant prognostic factor in the multivariate analysis. Only very few data are available on the potential clinical value of TLG in different cancer models. Xie et al reported on prognostic value of TLG in head and neck cancer for long term survival *(40),* while Cazaentre et al. demonstrated the usefulness of TLG for predicting response to radioimmunotherapy in lymphoma *(41)*. To date, the limited use of TV and TLG in clinical studies could be explained by the poor accuracy, robustness and reproducibility of available tumor delineation tools affecting the clinical value of resulting measurements. The fact that TLG was not an independent prognostic factor whereas TV alone was, suggests that the prognostic value of TLG mainly comes from the volume information and is impaired by the low prognostic value of SUV_{mean} measurements. In addition, the value of TLG might be reduced by a loss of information when combining the TV and the SUV_{mean} into one parameter by simple product, since large tumors with low uptake might result in the same TLG as small tumors with high uptake. Finally, the lack of partial volume effects (PVE) correction might also play a role in the reduced prognostic value of all SUV measurements as

well as the resulting TLG, since tumor volumes across the patients range from quite small and significantly affected by PVE (<2cm in diameter) to very large tumors for which PVE have smaller quantitative impact.

As expected, results concerning parameters dependent on the tumor delineation process were well correlated. On the other hand our results also highlighted the potential impact of differences between existing tumor delineation methods, with TV and TL being independent survival prognostic factors only when determined using FLAB. This approach has been previously shown to be both robust and accurate *(24,27)*. At present most commonly used methods are based on fixed or adaptive thresholds. Fixed thresholding has been demonstrated to be both inaccurate and non-robust *(23,24,27,39)* and was therefore not considered in this study.

Regarding the adaptive thresholding performance, results from one observer (T_{bckgrd}^{2}) were closer to these of FLAB compared to the other one (T_{bckgrd}^{1}), with however significant differences, as shown in figure 2B and table III. Neither TV(T_{bckgrd}^{1}) nor TV(T_{bckgrd}^{2}) were independent prognostic factors contrary to TV(FLAB). This can be explained by the behavior of adaptive thresholding (independently on the user) for several tumors. Most of the tumors exhibited simple shapes and homogeneous tracer uptake. However some were more complex and exhibited higher heterogeneity, or were small (<2-3cm) with low contrast. Adaptive thresholding has been demonstrated to provide unsatisfactory delineation for such cases *(24)*, because its final threshold is based on the ratio between an isocontour at 70% of the maximum and the background ROI. Such isocontour tends to overestimate (respectively underestimate) the actual value of the entire tumor for heterogeneous uptake (respectively small tumors will low contrast).

Hence the adaptive thresholding led to significant under-evaluation of larger heterogeneous tumors in our study, e.g. a patient with a survival of 6 months had a TV defined by FLAB of almost 97cm^3, whereas TV(T_{bckgrd}^1) and TV(T_{bckgrd}^2) were 38cm^3 (-61%) and 50cm^3 (-50%) respectively, clearly missing parts of the tumor. On the other hand, the dependency on the background ROI is higher regarding small tumors with low contrast, e.g. for a patient with 21 months survival, TV(FLAB) was 5.8cm^3, whereas TV(T_{bckgrd}^1) and TV(T_{bckgrd}^2) were 1.9cm^3 (-67%) and 26.9cm^3 (+364%) respectively. Several patients were therefore incorporated in the wrong survival curve, mostly patients with large volumes that were underestimated by the adaptive threshold.

In addition, adaptive thresholding was found to be highly user dependent, since we observed a bias up to 30% between the two users measuring TV, the agreement interval being too large for clinical applications (-110% to +45%). This seems to be in agreement with results concerning the level of reproducibility in measuring functional TV from ^{18}F-FDG imaging which can vary from 21% to 90% using automatic and threshold-based approaches respectively *(42)*. If advanced segmentation algorithms are not available, the use of adaptive thresholding approaches should be preferred to manual or fixed threshold-based delineation. Automated background ROI determination could reduce the inter observer variability observed in this work.

The limits of this study are its retrospective nature and the limited number of patients. Our results need to be confirmed through a prospective study on a larger cohort of patients. It is finally worth noting that overall survival might have been affected by other factors such as subsequent treatment for patients who relapsed, although this should have minor impact to the results of this study since it applies to

all considered parameters. Other outcome measures such as progression-free survival were not investigated in this study.

5. Conclusion

Our results suggest that the functional tumor volume followed by length has additional value compared to commonly-used SUV measurements (SUV_{max}, SUV_{peak}, SUV_{mean}) for prognosis in patients with locally advanced esophageal cancer treated with exclusive concomitant radio-chemotherapy. Both parameters were significant prognostic factors of overall survival, independently of the approach used to delineate the tumors. However, only the automatic FLAB algorithm allowed TV and TL to be independent prognostic factors of survival in a multivariate analysis that included standard tumor staging. In addition the total lesion glycolysis index was a statistically significant, but not independent, prognostic factor irrespective of the delineation algorithm used. Our findings confirm the potential value of ^{18}F-FDG PET to give a useful orientation for patient management purpose in esophageal cancer but they also highlight the influence of the methodology used on the degree of pertinence of these PET image derived parameters of interest as their accuracy and their clinical significance increase if they are computed using more reliable and robust tumor segmentation methods.

Conflict of interest: The authors declare that they have no conflict of interest.

References

(1) Falk GW. Risk factors for esophageal cancer development. *Surg Oncol Clin N Am*. 2009;18(3):469-85.

(2) Flamen P, Lerut A, Van Cutsem E, et al. Utility of positron emission tomography for the staging of patients with potentially operable esophageal carcinoma. *J Clin Oncol.* 2000;18:3202-10.

(3) Heeren PA, Jager PL, Bongaerts F, Van Dullemen H, Sluiter W, Plukker JT. Detection of distant metastases in esophageal cancer with (18)F-FDG PET. *J Nucl Med.* 2004;45:980-7.

(4) Van Vliet EP, Heijenbrok-Kal MH, Hunink MG, Kuipers EJ, Siersema PD. Staging investigations for esophageal cancer: a meta-analysis. *Br J Cancer.* 2008;98(3):547-57.

(5) Kim TJ, Kim HY, Lee KW, Kim MS. Multimodality assessment of esophageal cancer: preoperative staging and monitoring of response to therapy. *Radiographics.* 2009;29(2):403-2.

(6) Chuang HH, Macapinlac HA. The evolving role of PET-CT in the management of esophageal cancer. *Q J Nucl Med Mol Imaging.* 2009;53(2):201-9.

(7) Use of PET and PET/CT for radiation therapy planning: IAEA expert report 2006-2007. *Radiother Oncol.* 2009;91(1):85-94.

(8) MacManus M, Nestle U, Rosenzweig KE, et al. PET-based treatment planning in radiotherapy: a new standard?. *J Nucl Med.* 2007;48(S1):68S-77S.

(9) Grégoire V, Haustermans K, Geets X, et al. The value of PET/CT in gross tumor volume delineation in lung and esophagus cancer. *Int J Radiat Oncol Biol Phys.* 2004;60(S):S536-S537.

(10) Choi JY, Jang HY, Shim YM, et al. 18F-FDG PET in patients with esophageal squamous cell carcinoma undergoing curative surgery: prognostic implications. *J Nucl Med*. 2004;45(11):1843-50.

(11) Mamede M, Abreu-E-Lima P, Oliva MR, et al. FDG-PET/CT tumor segmentation-derived indices of metabolic activity to assess response to neoadjuvant therapy and progression-free

survival in esophageal cancer: correlation with histopathology results. *Am J Clin Oncol.* 2007;30(4):377-88.

(12) Hyun SH, Choi JY, Shim YM, et al. Prognostic Value of Metabolic Tumor Volume Measured by 18F-Fluorodeoxyglucose Positron Emission Tomography in Patients with Esophageal Carcinoma. *Ann Surg Oncol.* 2010;17:115-122.

(13) Hong D, Lunagomez S, Kim EE, et al. Value of baseline positron emission tomography for predicting overall survival in patient with nonmetastatic esophageal or gastroesophageal junction carcinoma. *Cancer.* 2005;104:1620-6.

(14) Blackstock AW, Farmer MR, Lovato J, et al. A prospective evaluation of the impact of 18-F-fluoro-deoxy-D-glucose positron emission tomography staging on survival for patients with locally advanced esophageal cancer. *Int J Radiat Oncol Biol Phys.* 2006; 64:455-60.

(15) Van Westreenen HL, Plukker JT, Cobben DC, Verhoogt CJ, Groen H, Jager PL. Prognostic value of the standardized uptake value in esophageal cancer. *Am J Roentgenol.* 2005;185(2):436-40.

(16) Omloo JM, Sloof GW, Boellaard R, et al. Importance of fluorodeoxyglucose-positron emission tomography (FDG-PET) and endoscopic ultrasonography parameters in predicting survival following surgery for esophageal cancer. *Endoscopy.* 2008;40(6):464-71.

(17) Kato H, Nakajima M, Sohda M, et al. The clinical application of 18FDG positron emission tomography to predict survival in patients with operable esophageal cancer. *Cancer.* 2009:3196-3203.

(18) Cheze-Le Rest C, Metges JP, Teyton P, et al. Prognostic value of initial fluorodeoxyglucose-PET in esophageal cancer: a prospective study. *Nucl Med Commun.* 2008;29:628-635.

(19) Cerfolio RJ, Bryant AS. Maximum standardized uptake values on positron emission tomography of esophageal cancer predicts stage, tumor biology, and survival. *Ann Thorac Surg.* 2006;82:391-395.

(20) Rizk N, Downey RJ, Akhurst T, et al. Preoperative 18FDG positron emission tomography standardized uptake values predict survival after esophageal adenocarcinoma resection. *Ann Thorac Surg.* 2006;81:1076-1081.

(21) Yendamuri S, Swisher SG, Correa AM, et al. Esophageal Tumor Length Is Independently Associated with Long-term Survival. *Cancer.* 2009;115:508-16.

(22) Roedl JB, Harisinghani MG, Colen RR, et al. Assessment of treatment response and recurrence in esophageal carcinoma based on tumor length and standardized uptake value on positron emission tomography-computed tomography. *Ann Thorac Surg.* 2008;86(4):1131-8.

(23) Nestle U, Kremp S, Schaefer-Schuler A, et al. Comparison of Different Methods for Delineation of 18F-FDG PET-Positive Tissue for Target Volume Definition in Radiotherapy of Patients with Non-Small Cell Lung Cancer. *J Nucl Med.* 2005;46(8):1342-8.

(24) Hatt M, Cheze-le Rest C, Descourt P, et al. Accurate automatic delineation of heterogeneous functional volumes in positron emission tomography for oncology applications. *Int J Radiat Oncol Biol Phys*. 2010;77:301-308.

(25) Larson SM, Erdi Y, Akhurst T, et al. Tumor treatment response based on visual and quantitative changes in global tumor glycolysis using PET-FDG imaging. The visual response score and the change in total lesion glycolysis. *Clin Positron Imaging*. 1999;2:159–71

(26) Velasquez LM, et al. Repeatability of 18F-FDG PET in a Multicenter Phase I Study of Patients with Advanced Gastrointestinal Malignancies, *J Nucl Med.* 2009;50(10):1646-1654.

(27) Hatt M, Turzo A, Roux C, et al. A fuzzy Bayesian locally adaptive segmentation approach for volume determination in PET. *IEEE Trans Med Im.* 2009;28(6):881-893.

(28) Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986;1:307-310.

(29) Kaplan E, Meyer P. Non parametric estimation from incomplete observations. *J Am Stat Assoc*. 1958;53:457-481.

(30) Metz CE. Basic principles of ROC analysis. *Semin Nucl Med*. 1978;8(4):283-98.

(31) Greene FL, Page DL, Fleming ID et al. AJCC cancer staging manual, ed 6, new York, Springer, 2002

(32) Cox DR. Regression Models and Life Table. *Journal of the Royal Statistical Society Series B.* 1972:34(2):187-220.

(33) van Heijl M, Omloo JM, van Berge Henegouwen MI, van Lanschot JJ, Sloof GW, Boellaard R. Influence of ROI definition, partial volume correction and SUV normalization on SUV-survival correlation in oesophageal cancer. Nucl Med Commun. 2010 Jul;31(7):652-8.

(34) Himeno S, Yasuda S, Shimada H, Tajima T, Makuuchi H. Evaluation of esophageal cancer by positron emission tomography. *Jpn J Clin Oncol.* 2002;32:340-6.

(35) Fukunaga T, Okazumi S, Koide Y, Isono K, Imazeki K. Evaluation of esophageal cancers using fluorine-18-fluorodeoxyglucose PET. *J Nucl Med*. 1998;39:1002-7.

(36) Taylor MD, Smith PW, Brix WK, et al. Correlations between selected tumor markers and fluorodeoxyglucose maximal standardized uptake values in esophageal cancer. *Eur J Cardiothorac Surg.* 2009;35:699-705.

(37) Swisher S, Erasmus J, Maish M, et al. 2Fluoro-2-deoxy-D-glucose positron emission tomography imaging is predictive of pathologic response and survival after preoperative chemoradiation in patients with esophageal carcinoma. *Cancer.* 2004;101:1776-1785.

(38) Zhong X, Yu J, Zhang B, et al. Using 18F-fluorodeoxyglucose positron emission tomography to estimate the length of gross tumor in patients with squamous cell carcinoma of the esophagus. *Int J Radiat Oncol Biol Phys.* 2009;73(1):136-141.

(39) Tylski P, Stute S, Grotus N, et al. Comparative assessment of methods for estimating tumor volume and standardized uptake value in (18)F-FDG PET. *J Nucl Med*. 2010;51(2):268-76.

(40) Xie P, Yue JB, Zhao HX, et al. Prognostic value of (18)F-FDG PET-CT metabolic index for nasopharyngeal carcinoma. *J Cancer Res Clin Oncol.* 2010;136(6):883-889.

(41) Cazaentre T, Morschhauser F, Vermandel M, et al. Pre-therapy 18F-FDG PET quantitative parameters help in predicting the response to radioimmunotherapy in non-Hodgkin lymphoma. *Eur J Nucl Med Mol Img.* 2010;37(3):494-504.

(42) M. Hatt, C. Cheze Le Rest, E.O. Aboagye, et al. Reproducibility of 18F-FDG and 18F-FLT PET tumor volume measurements *J Nucl Med.* 2010; 51(9):1368-1376.

Figures captions

Fig.1: Illustration of differences in tumor delineation depending on the methodology for (A) a small (<8cm^3) and low contrast (approximately 2:1) and (B) a larger (30cm^3) and higher contrast (approximately 7:1) tumors.

Fig.2: Bland-Altman analysis of differences between (A) T_{bckgrd}^1 and T_{bckgrd}^2 and (B) T_{bckgrd} and FLAB, for each parameter (TL, TV, SUV_{mean}, TLG). Grey columns and error bars represent the mean differences (bias) and associated standard deviation respectively. Bold arrows up and down denote upper and lower limits respectively. 95% confidence intervals for each are given in table III.

Fig.3: Kaplan-Meier survival curves obtained using (A) SUV_{max}, TV measured by (B) FLAB and (C) T_{bckgrd}^1, and (D) defining 3 groups using TV measured by FLAB.

Fig.4: ^{18}F-FDG PET images (axial, coronal and sagittal views from top to bottom) of patients with a (a) small tumor (11cm^3, 54 months survival), (b) medium size tumor (22cm^3, 18 months survival) and (c) larger tumor (92cm^3, 5 months survival).

Table captions

Table I: Patient demographic and clinical characteristics

Table II: Parameters definition and statistics.

Table III: Bland-Altman analysis results comparing delineation strategies for all parameters.

Table IV: Parameters with significant prognostic value after univariate analysis.

Table V: Prognostic value of all parameters.

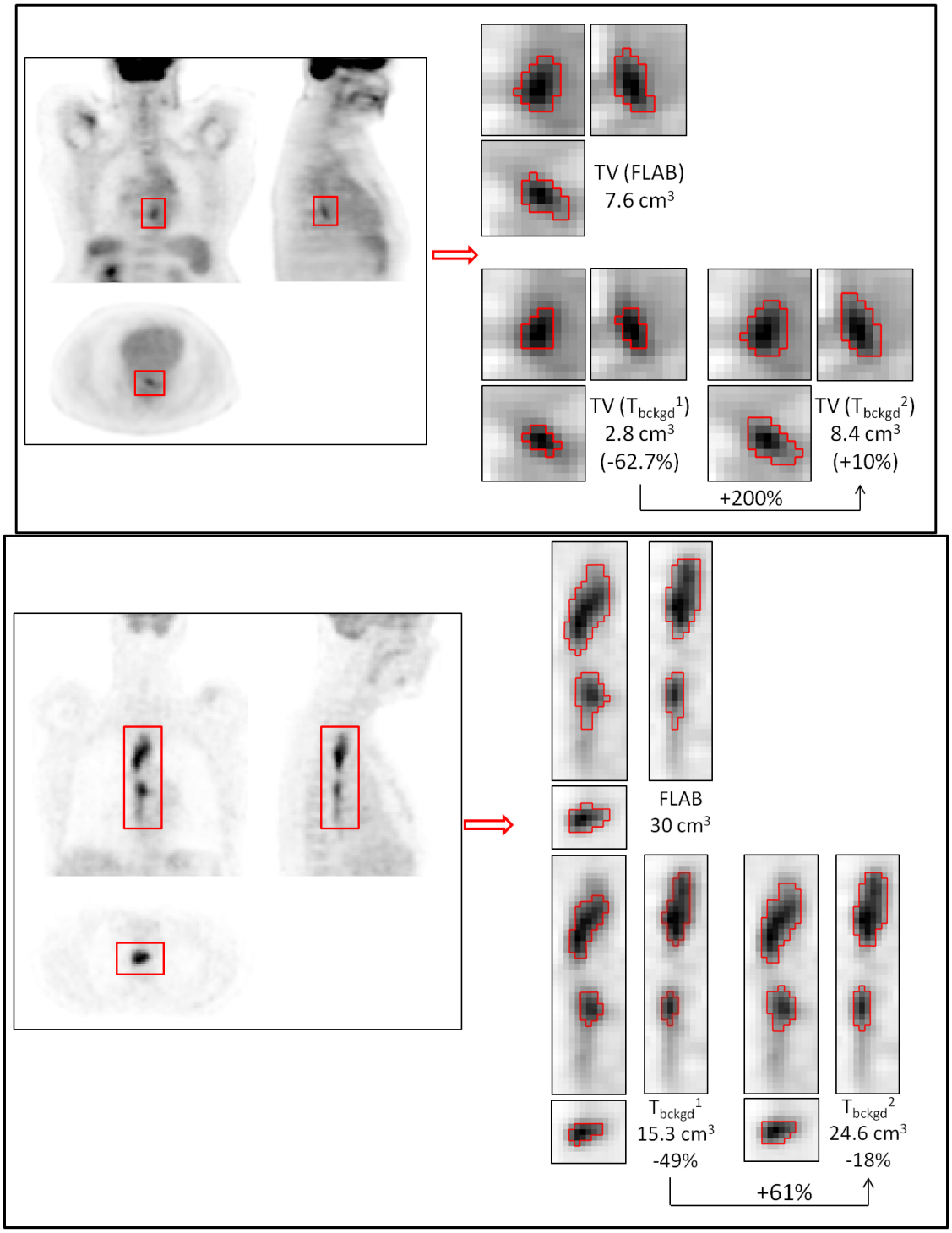

TV (FLAB)
7.6 cm3
TV (Tbckgd1)
2.8 cm3
(-62.7%)
TV (Tbckgd2)
8.4 cm3
(+10%)
+200%
FLAB
30 cm3
Tbckgd1
15.3 cm3
-49%
Tbckgd2
24.6 cm3
-18%
+61%

Figure 1

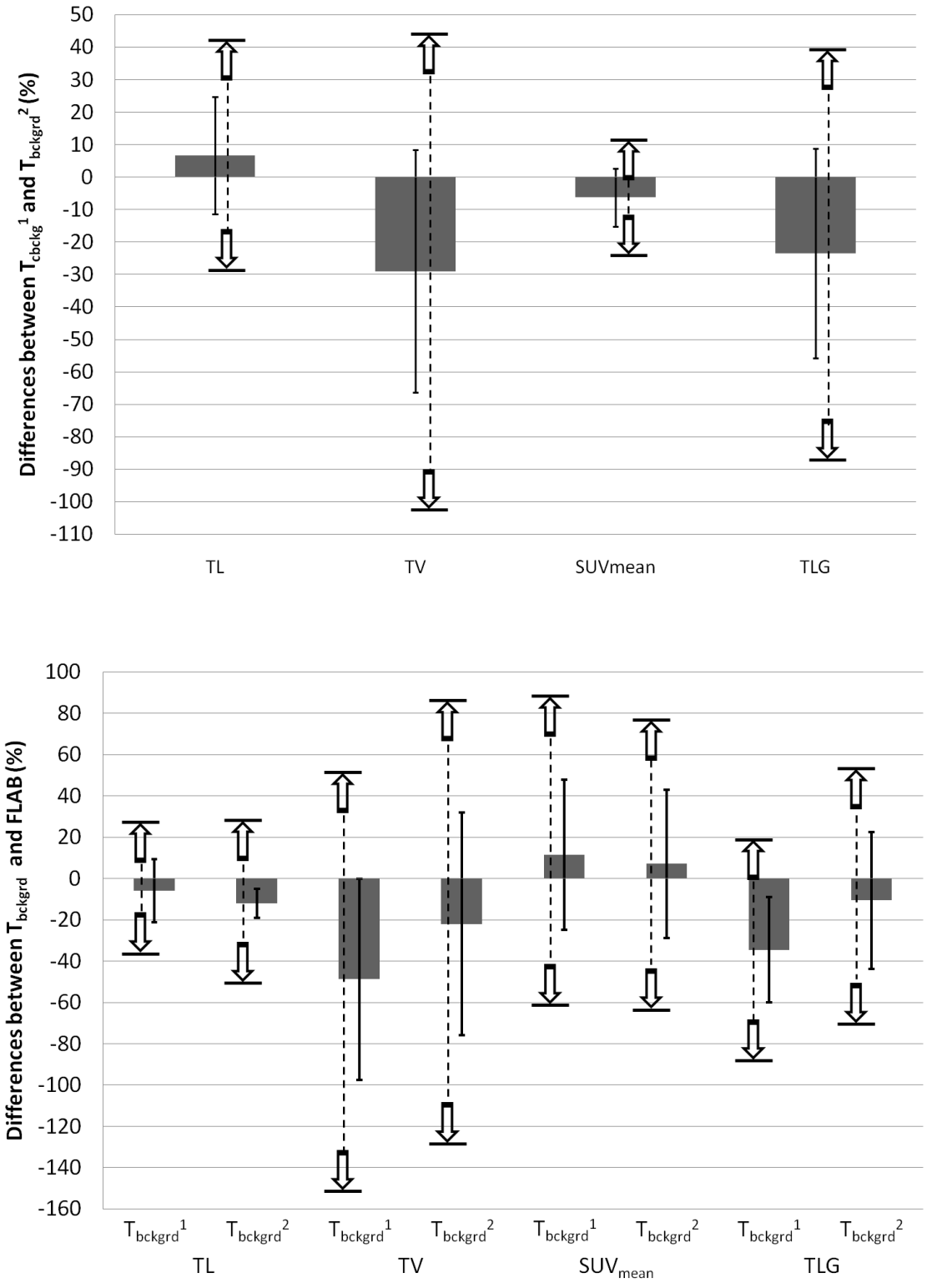
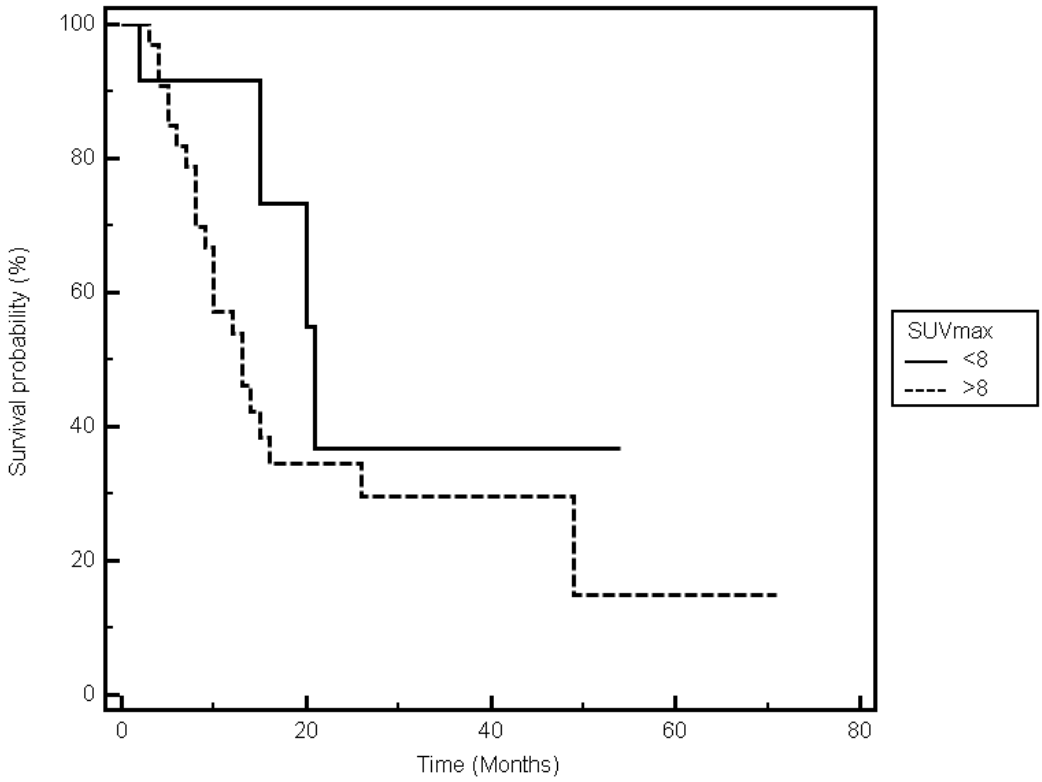
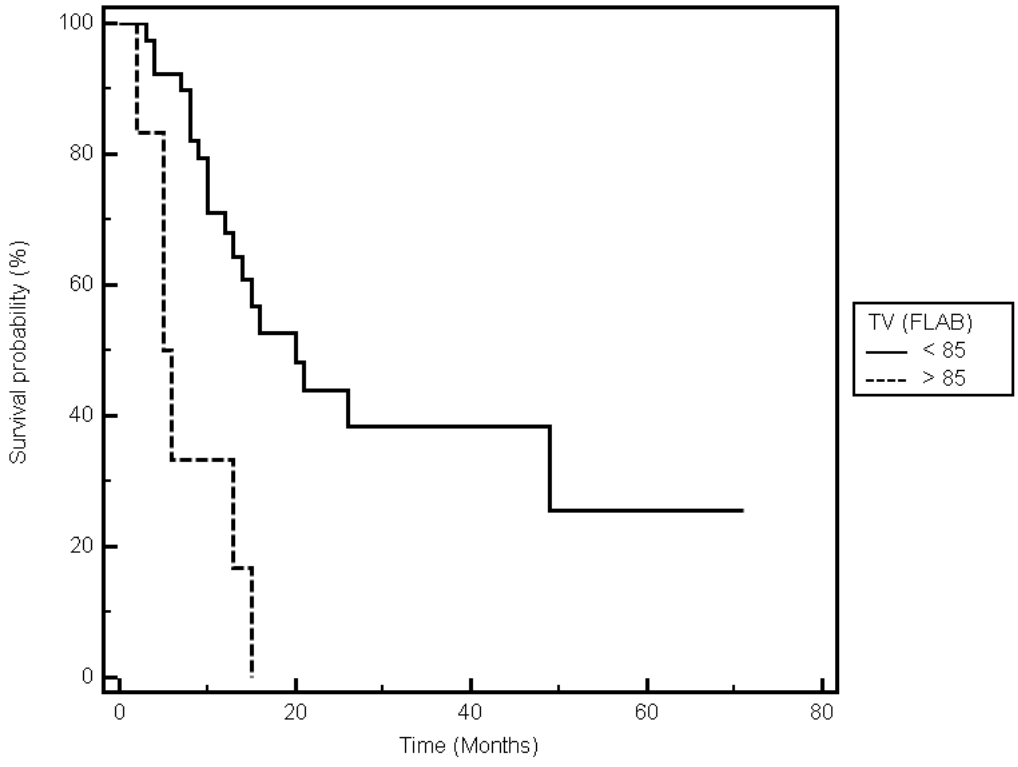
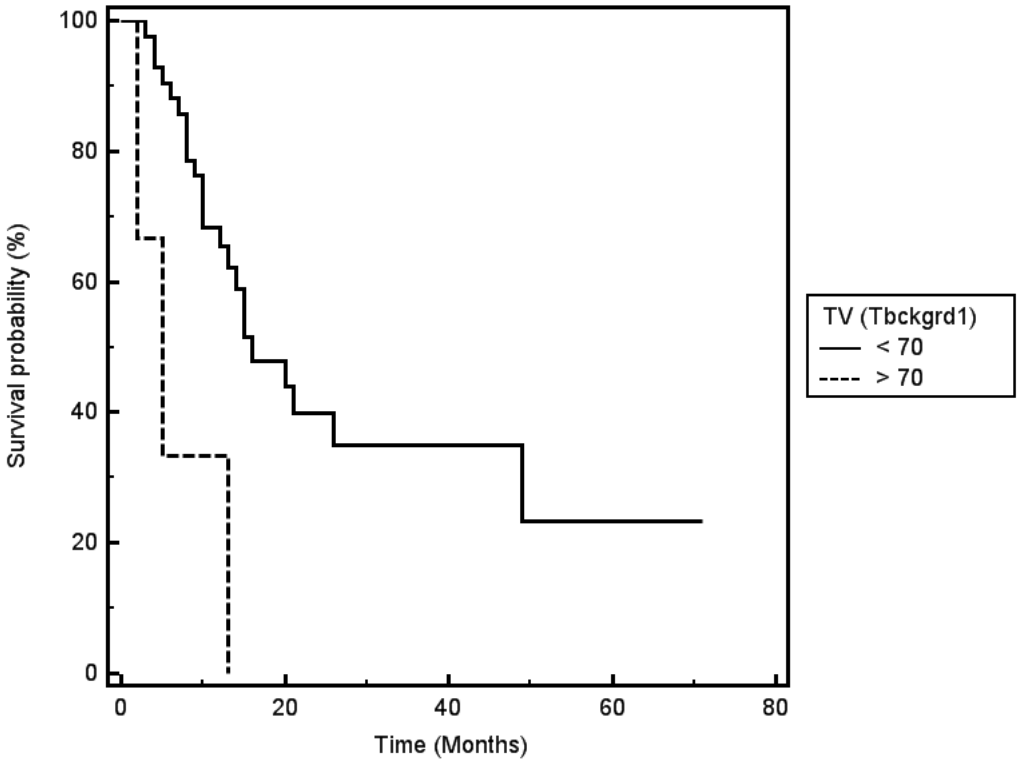

Differences between T_{cbckg}^{1} and T_{bckgrd}^{2} (%)
50
40
30
20
10
0
-10
-20
-30
-40
-50
-60
-70
-80
-90
-100
-110
TL
TV
SUVmean
TLG
Differences between T_{bckgrd} and FLAB (%)
100
80
60
40
20
0
-20
-40
-60
-80
-100
-120
-140
-160
T_{bckgrd}^{1}
T_{bckgrd}^{2}
TL
TV
SUV_{mean}
TLG

Figure 2


100
80
60
40
20
0
Survival probability (%)
0
20
40
60
80
Time (Months)
SUVmax
<8
>8


100
80
60
40
20
0
Survival probability (%)
0
20
40
60
80
Time (Months)
TV (FLAB)
< 85
> 85


Survival probability (%)
100
80
60
40
20
0
0
20
40
60
80
Time (Months)
TV (Tbckgrd1)
< 70
> 70

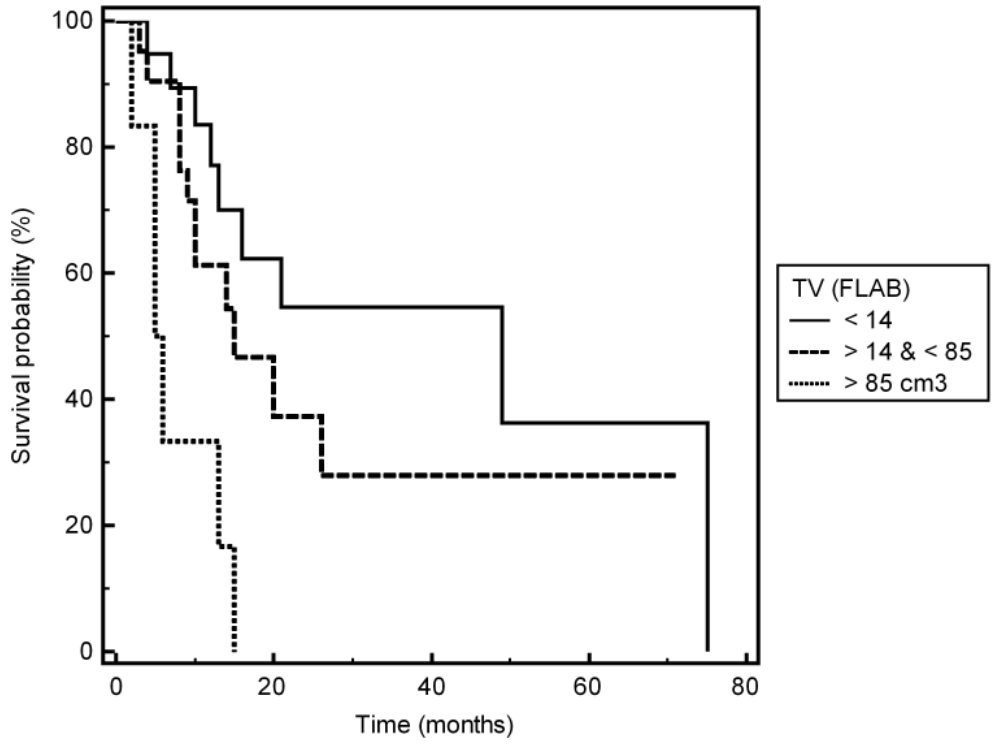

Survival probability (%)
100
80
60
40
20
0
0
20
40
60
80
Time (months)
TV (FLAB)
< 14
> 14 & < 85
> 85 cm3

Figure 3

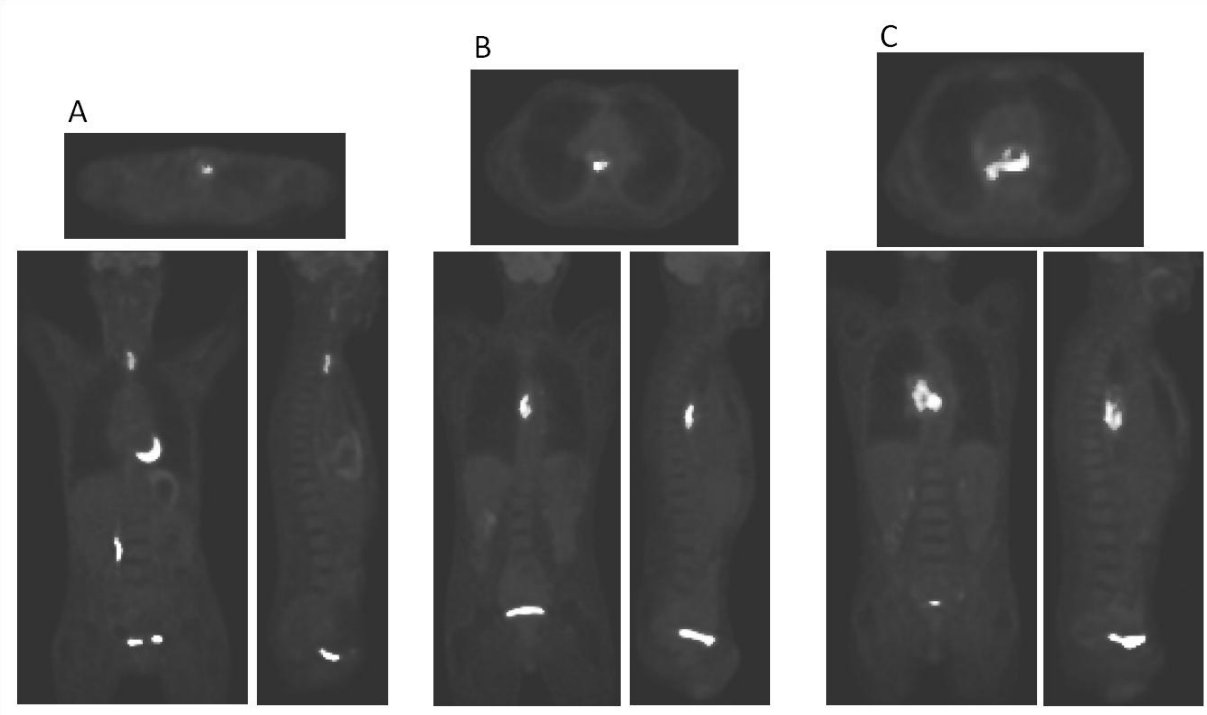


Figure 4

Parameter	# of patients (%)
Gender	
Male	41(91)
Female	4(9)
Age	
Range	45-84
Median	68
Site	
Upper esophagus	11(24)
Middle esophagus	17(38)
Lower esophagus	17(38)
Histology type	
Adenocarcinoma	12(27)
Squamous cell carcinoma	33(73)
Histology differentiation	
Well differentiated	12(27)
Moderately differentiated	11(24)
Poorly differentiated	4(9)
Unknown	18(40)
TNM Stage	
T1	6(13)
T2	7(16)
T3	22(49)
T4	10(22)
N0	18(40)
N1	27(60)
M0	29(64)
M1	16(36)
AJCC Stage	
I	3(7)
IIA	7(16)
IIB	5(11)
III	14(31)
IVa	5(11)
IVb	11(24)

Table I

Definition			Notation	Mean±SD	Range
Highest SUV within the tumor			SUV_{max}	10±3.8	2.2 – 19.7
Mean of SUV_{max} and its 26 neighbors			SUV_{peak}	8.2±3.3	1.8 – 16.1
Mean SUV of tumor defined by	Adaptive threshold	1st user	$SUV_{mean}(T_{bckgrd}^{1})$	6.6±2.6	1.8 – 13.7
		2nd user	$SUV_{mean}(T_{bckgrd}^{2})$	6.2±2.7	1.6 – 13.8
	FLAB		SUV_{mean}(FLAB)	6.0±2.4	1.7 – 13.2
Functional tumor volume defined by	Adaptive threshold	1st user	$TV(T_{bckgrd}^{1})$	22.6±23.8	1.8 – 92.0
		2nd user	$TV(T_{bckgrd}^{2})$	29.2±29.7	2.4 – 133.9
	FLAB		TV(FLAB)	36.3±33.7	3.0 – 139.7
Functional tumor length defined by	Adaptive threshold	1st user	$TL(T_{bckgrd}^{1})$	5.9±3.0	1.6 – 15.6
		2nd user	$TL(T_{bckgrd}^{2})$	5.6±2.9	1.6 – 14.4
	FLAB		TL(FLAB)	6.2±2.9	2.0 – 15.6
$SUV_{mean}(T_{bckgrd}^{1})xTV(T_{bckgrd}^{1})$(g)			$TLG(T_{bckgrd}^{1})$	165.4±182.7	3.2 – 759.7
$SUV_{mean}(T_{bckgrd}^{2})xTV(T_{bckgrd}^{2})$(g)			$TLG(T_{bckgrd}^{2})$	198.8±209.4	6.9 – 921.3
SUV_{mean}(FLAB)xTV(FLAB)(g)			TLG(FLAB)	221.6±225.8	5.3 – 882.7

Table II

Parameter	% difference between T_{bckgrd}^1 and T_{bckgrd}^2					
	Mean±SD	95%CI of mean	LL	95%CI of LL	UL	95%CI of UL
TL	6.7±18	1.3 to 12.1	-28.6	-37.9 to -19.3	41.9	32.6 to 51.2
TV	-29±37.3	-40.2 to -17.8	-102	-121.3 to -82.8	44.1	24.8 to 63.4
SUV_{mean}	-6.3±9	-9 to -3.6	-23.9	-28.5 to -19.3	11.2	6.6 to 15.8
TLG	-23.5±32.3	-33.2 to -13.8	-86.8	-103.5 to -70.1	39.7	23 to 56.4

Parameter		% difference between T_{bckgrd} and FLAB					
		Mean±SD	95%CI of mean	LL	95%CI of LL	UL	95%CI of UL
TL	T_{bckgrd}^1	-5.9 ± 15.3	-10.4 to -1.4	-35.8	-43.6 to -28	24	16.2 to 31.8
	T_{bckgrd}^2	-12 ± 7	-18.3 to 7.1	-49.4	-59 to -39.9	24.1	14.5 to 33.6
TV	T_{bckgrd}^1	-48.8 ± 48.8	-63.3 to -34.3	-144.5	-169.5 to -120	46.9	21.9 to 71.9
	T_{bckgrd}^2	-22 ± 53.9	-38.1 to -6.0	-127.7	-155.3 to -100	83.6	56.1 to 111.2
SUV_{mean}	T_{bckgrd}^1	11.5 ± 36.2	0.7 to 22.2	-59.5	-78 to -41	82.4	63.8 to 100.9
	T_{bckgrd}^2	7.1 ± 35.8	-3.6 to 17.7	-63.1	-81.4 to -44.8	77.2	58.9 to 95.5
TLG	T_{bckgrd}^1	-34.5 ± 25.6	-42 to -26.9	-84.6	-97.6 to -71.5	15.7	2.6 to 28.7
	T_{bckgrd}^2	-10.6 ± 33.2	-20.4 to -0.7	-75.6	-92.5 to -58.6	54.4	37.5 to 71.4

SD: Standard Deviation. CI: Confidence Interval. UL: Upper Limit. LL: Lower Limit.

Table III

Parameter	HR	HR 95%CI	*P*	Median survival (months)
AJCC stage	0.281	0.090 – 0.881	0.0008	26*vs*8
M stage	0.402	0.172 – 0.940	0.01	26*vs*12
TL(T_{bckgrd}^{1})	0.318	0.133 – 0.761	0.02	21*vs*10
TL(T_{bckgrd}^{2})	0.393	0.164 – 0.939	0.04	21*vs*10
TL(FLAB)	0.163	0.052 – 0.510	0.01	21*vs*10
TV(T_{bckgrd}^{1})	0.212	0.020 – 2.280	0.004	16*vs*5
	N/A	N/A	0.02	21*vs*10*vs*9
TV(T_{bckgrd}^{2})	0.212	0.020 – 2.280	0.004	16*vs*5
	N/A	N/A	0.004	49*vs*14*vs*5
TV(FLAB)	0.236	0.050 – 0.909	0.0005	20*vs*5.5
	N/A	N/A	0.002	49*vs*15*vs*5.5
TLG(T_{bckgrd}^{1})	0.217	0.064 – 0.735	0.007	20*vs*8
TLG(T_{bckgrd}^{2})	0.202	0.063 – 0.645	0.01	20*vs*10
TLG(FLAB)	0.337	0.147 – 0.772	0.02	21*vs*10

HR: Hazard Ratio. CI: Confidence Interval

Table IV

Variable	***Significant prognostic factor in univariate analysis***	***Significant independent prognostic factor in multivariate analysis***
Age	No	-
Gender	No	-
Histology type	No	-
AJCC stage	**Yes**	**Yes**
T	No	-
N	No	-
M	**Yes**	No
SUV_{max}	No	-
SUV_{peak}	No	-
$SUV_{mean}(T_{bckgrd}^{1})$	No	-
$SUV_{mean}(T_{bckgrd}^{2})$	No	-
SUV_{mean}(FLAB)	No	-
$TL(T_{bckgrd}^{1})$	**Yes**	No
$TL(T_{bckgrd}^{2})$	**Yes**	No
TL(FLAB)	**Yes**	**Yes**
$TV(T_{bckgrd}^{1})$	**Yes**	No
$TV(T_{bckgrd}^{2})$	**Yes**	No
TV(FLAB)	**Yes**	**Yes**
$TLG(T_{bckgrd}^{1})$	**Yes**	No
$TLG(T_{bckgrd}^{2})$	**Yes**	No
TLG(FLAB)	**Yes**	No

Table V